# DFG-Projekt: Coli-conc
## Das Mapping-Tool „Cocoda“

*Uma Balakrishnan, Verbundzentrale des Gemeinsamen Bibliotheksverbunds*

**Zusammenfassung:**
Der Beitrag beschreibt das Projekt Coli-conc, das von der Verbundzentrale des Gemeinsamen Bibliotheksverbunds (GBV) betreut wird. Ziel ist die Entwicklung einer Infrastruktur für den Austausch, die Erstellung und die Wartung von Konkordanzen zwischen bibliothekarischen Wissensorganisationssystemen.

**Summary:**
This article describes the project Coli-conc, overseen by the Head Office of the Gemeinsamer Bibliotheksverbund (GBV). The aim is to develop an infrastructure for the exchange, the creation and maintenance of concordances between different knowledge organization systems for libraries.

**Zitierfähiger Link (DOI):** http://dx.doi.org/10.5282/o-bib/2016H1S11-16
**Schlagwörter:** Klassifikation; Konkordanz

Die Heterogenität der sowohl im deutschsprachigen Raum als auch weltweit angewendeten Klassifikationssysteme stellt eine Barriere im Informationsaustausch dar. Dies führt vor allem bei der Suche von Ressourcen in einem unterschiedlich erschlossenen Datenpool zu einem restriktiven Retrieval für die Endnutzerinnen und -nutzer. Die Aufgabe von Konkordanzen liegt in der Überwindung dieser Einschränkungen. Obwohl in den letzten Jahren die Nutzung und Verbreitung von Wissensorganisationssystemen (KOS) zur Erschließung von Dokumenten und Daten deutlich zugenommen hat, sind Konkordanzen zwischen verschiedenen KOS nur sehr begrenzt verfügbar.

Die intellektuelle Erstellung von Konkordanzen ist jedoch sehr aufwendig und stößt angesichts der Tiefe, Größe und kontrastiven Natur verschiedener KOS insbesondere bei fein gegliederten Klassifikationssystemen an praktische Grenzen. Automatische Verfahren zur Erstellung von Konkordanzen sind zwar möglich und Gegenstand der Forschung,[1] in der Regel aber unvollständig und eher für Mapping-Vorschläge geeignet.[2]

1 Vgl. „Ontology Alignment Evaluation Initiative (OAEI). Library Track (2012–2014),“ jeweils zuletzt geprüft am 05.01.2016, http://oaei.ontologymatching.org/2012/library/, http://oaei.ontologymatching.org/2014/library/, sowie Magnus Pfeffer, „Automatic creation of mappings between classification systems” (Vortrag auf der European Conference on Data Analysis, Luxemburg, 10. Juli 2013, zuletzt peprüft am 05.02.2016, http://de.slideshare.net/MagnusPfeffer/pfeffer-automatic-mapping.

2 Vgl. Boris Lauser et al., „Comparing human and automatic thesaurus mapping approaches in the agricultural domain“ (Vortrag auf der 10th International Conference on Dublin Core and Metadata Applications, 2008), zuletzt geprüft am 05.01.2016, http://arxiv.org/abs/0808.2246, und Ulrike Reiner, „Automatische DDC-Klassifizierung: Bibliografische Titeldatensätze der Deutschen Nationalbibliografie,“ *Dialog mit Bibliotheken* 22, Nr. 1 (2010): 23–29, zuletzt geprüft am 05.02.2016, http://nbn-resolving.de/urn:nbn:de:101-2011012860.

*Tabelle 1: Klassifikationssysteme im deutschsprachigen Raum*

| Classification systems in German speaking regions | |
|---|---|
| **Universal Classification Systems** | **No. of classes** |
| UDC (Universal Decimal Classification ) | ca. 65.000 classes (English version) |
| DDC (Dewey Decimal Classification) | over 44.000 classes with 10 main classes |
| RVK (Regensburg Classification) | 850.000 classes with 33 main classes |
| BC (Basic Classification) | 2100 classes with 89 main classess |
| LCC (Library of Congress Classification) | 21 main classes |
| **Subject classification** | **No. of classes** |
| DDC-Sachgruppen der DNB | 10 main classes with 94 subclasses |
| MSC (Mathematics Subject Classification) | 87 main classes |
| PACS (Physics and Astronomy Classification Scheme) | 10 main classes |
| FKDigBib (Subject classification for digital library) | 10 main classes |
| KfM (Classification for music library) | ca. 800 classes |
| **Subject Classification at the Universities** | **No. of classes** |
| TUM-classification (Science and technologyl classification of the TU Munic) | 52 classes each with 999 notations |
| Subject classification of the University library Duesseldorf | 45 classes |
| Bremer classification of the State and University library Bremen | ca. 57 main classes |
| GOK (Goettingen Online Classification) | ca. 33 main classes |
| Standard-Thesaurus Wirtschaft von der ZWB | 6.000 Terms and notations |
| Subject classification University library Trier | 36 main classes |
| Technical University Dortmund | 28 main classes |
| University library Paderborn | 26 main classes |
| University library Marburg | 35 main classes |
| University library Bonn | 24 main classes |
| University library Heidelberg | 22 main classes |
| Subject classification and nomenclature of individual languages Library of the Institute of General Linguistics at the Uni Münster | 23 main classes |
| **Subject Classification at the public libraries** | **No. of classes** |
| SEB (Scheme for protestant libraries) | |
| SKB-E (Scheme for catholic public libraries) | |
| KfKJ (Scheme for children and youth libraries) | Less than 1.000 classes |
| ASB (General classification for public libraries) | ca. 2.200 classes with 23 main classes |
| ÖSÖB (Austrian classification for public libarries) | |
| SfB (Classification for libraries) | ca 14.400 classes with 30 main classes |
| KAB (Classification for general libraries) | ca. 2.700 classes |
| SSD (Classfication of the city library Duisburg) | |
| ESSB (Single classfication for South Tyrolean) | 16 main classes |

Die geringe Verfügbarkeit von Konkordanzen hat zur Folge, dass sich trotz Bemühungen zur Standardisierung[3] weder Austauschformate noch Verfahren für die nachhaltige Pflege und Bereitstellung durchgesetzt haben. Vor allem im Bereich bibliothekarischer Klassifikationen sind keine Werkzeuge

3 Vgl. Stefan Keil, „Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben," *Information – Wissenschaft und Praxis* 63, Nr. 1 (2012): 45–55, http://dx.doi.org/10.1515/iwp-2012-0004; „SKOS," zuletzt geprüft am 05.01.2015, http://openskos.org/; ISO, „ISO 25964-2: Information and documentation – Thesauri andinteroperability with other vocabularies. Part 2: Interoperability with other vocabularies"; Dagobert Soergel, „Conceptual foundations for semantic mapping and semantic search," in *Concepts in context*, hrsg. Felix Boteram et al. (Würzburg: Ergon, 2011), 13–35.

 

und etablierte Verfahren bekannt. Grundsätzlich lässt sich feststellen, dass es sowohl an einer Infrastruktur zur Bereitstellung und dem Austausch von Konkordanzen, als auch an Werkzeugen für deren Bearbeitung und qualitativen Bewertung mangelt.

Das Konkordanz-Projekt Coli-conc der Verbundzentrale des GBV, das vor Kurzem von der DFG für zwei Jahre bewilligt worden ist, hat deshalb zum Ziel, ein Tool zu entwickeln, das sowohl die intellektuelle Erstellung von Konkordanzen beschleunigt und vereinfacht, als auch deren Nutzung und Austausch vorantreibt, indem es als eine Plattform für Zusammenarbeit von Expertinnen und Experten dient und Konkordanzen für eine freie gemeinsame Nutzung zur Verfügung stellt. Zudem ist die Integration bereits vorhandener Mappings aus verschiedenen Projekten in die von dem Tool verwaltete Datenbank vorgesehen. Auch die Anbindung des Tools an die bibliothekarische Katalogisierungssoftware ist geplant, um eine nachhaltige Nutzung zu gewährleisten.

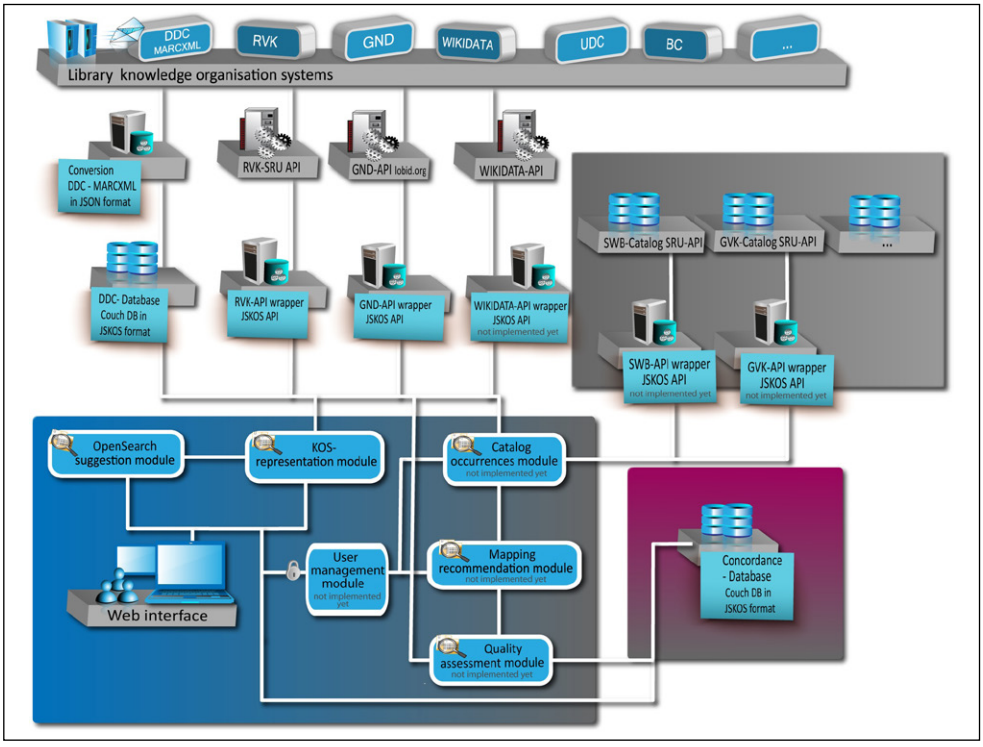


*Abb. 1: Modularer Aufbau des Tools.*

Alle Komponenten sollen über offene APIs miteinander verbunden werden, so dass sie als verteilte Infrastruktur für die Erstellung, den Austausch und die Wartung von Konkordanzen verwendet werden können. Derzeit ist nur die Anbindung an die GND (über lobid.org) und an die RVK (über API der Uni Regensburg) umgesetzt. Die DDC-Daten wurden uns von OCLC als XML-Dumps zur Verfügung gestellt und lokal gespeichert.

Zur gemeinsamen Nutzung verschiedener KOS und ihrer Mappings werden diese zunächst in ein einheitliches, mit SKOS kompatibles Datenformat konvertiert. Das im Rahmen des Projektes entwickelte Format JSKOS (Cocoda TR 2) bringt die Vorteile von RDF zur Datenaggregation in Linked-Data-Umgebungen ohne die volle Komplexität der RDF ins Spiel zu bringen und nutzt das JSON-Format für eine einfache Manipulation der Daten und deren Speicherung. Zusätzlich zu den in SKOS definierten Datenelementen ermöglicht JSKOS Aussagen über Mappings und Konkordanzen, wie z.B. *creator*, *degree of alignment* und *mapping methods*. Ein Skript zur Konvertierung zwischen SKOS und JSKOS wird derzeit entwickelt.[4]

Zur Speicherung der nach JSKOS konvertierten Daten werden derzeit verschiedene NoSQL-Datenbanken wie CouchDB und MongoDB evaluiert. Bereits jetzt werden die Mappings aller zur Verfügung stehenden Konkordanzen in einer Datenbank gespeichert und sind über die Projekthomepage abrufbar.[5]

Die Benutzeroberfläche des als Cocoda bezeichneten Tools zur Konkordanzerstellungsebene basiert auf mehreren Javascript-Modulen für das AngularJS-Framework. Zwei dieser Module sind bereits in einer Alpha-Version publiziert.

Das *KOS-Representation Module* (ng-skos) dient der Darstellung von Begriffen und Strukturen von KOS und Mappings, beispielsweise zur Navigation und Auswahl einer Notation.

Das *OpenSearch Suggestions Module* (ng-suggest) zeigt Vorschläge bei der Begriffssuche in einem KOS an.

Die weiteren auf der Abbildung 1 dargestellten Module sind noch zu implementieren:

Das *Mapping Recommendation Module* ist für die automatische Ermittlung von Mappingvorschlägen zuständig. Dies geschieht unter anderem durch

- die automatische Suche nach Benennungen, Begriffen und deren Synonymen im Ziel-KOS.
- die Evaluierung zusammen vorkommender Notationen und Begriffe verschiedener Normdaten in den Titeldatensätzen unterschiedlicher Kataloge.
- die Abfrage gespeicherter Konkordanzen in der VZG-Konkordanzdatenbank.
- die Einbeziehung von Ergebnissen einer manuellen Suche im Zielsystem.

Das *Quality-Assessment Module* ist verantwortlich für die Überprüfung

- der Korrektheit der Notationen und deren Benennungen.
- der Aktualität der Notationen und Begriffe.
- der Vollständigkeit der Konkordanz für ein Fachgebiet in einer Systematik.
- für die automatische Evaluierung der „Confidence rate“ der maschinell erzeugten Mappingvorschläge.

4 siehe https://github.com/gbv/skos2jskos.
5 https://coli-conc.gbv.de/

 

Das *User-Management Module* übernimmt die Verwaltung bzw. Authentifizierung von Bibliotheken und Nutzern, die das Tool nutzen bzw. ihre Mappings in der Konkordanzdatenbank speichern wollen.

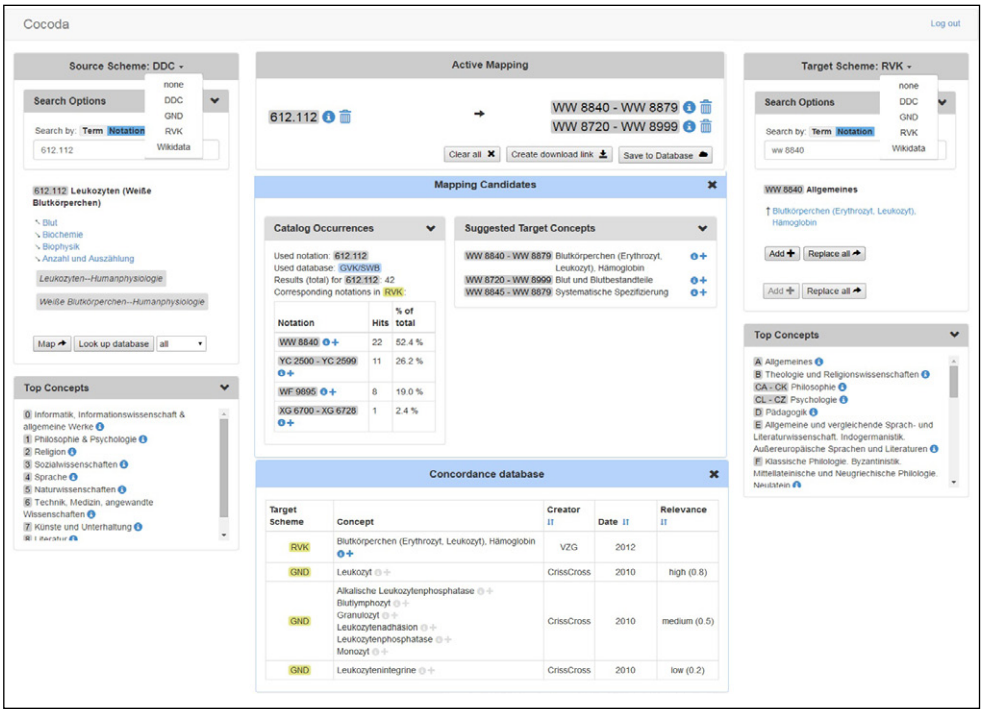


*Abb. 2: Die Benutzeroberfläche eines funktionsfähigen Prototypen des Konkordanz-Tools „Cocoda“.*

Der Schwerpunkt bei der Konkordanzerstellung liegt im Rahmen des Projektes zunächst auf den im deutschsprachigen Raum weit verbreiteten Klassifikationssystemen wie DDC, RVK und BK, deshalb wird zunächst auf die Konkordanzerstellung zwischen DDC und RVK eingegangen.

Das dashboardartige Darstellungskonzept stellt alle notwendigen Informationen für die Erstellung der Konkordanzen auf einem Bildschirm bereit. Gegliedert wird die Tooloberfläche in drei Teile. Dabei sind links und rechts die Auswahlfenster der aufeinander abzubildenden Systeme und in der Mitte die Fenster des Ergebnisbereiches platziert. Die Suchoptionsfenster bieten die Möglichkeit, die Suche anhand der Begriffs- oder Notationsangaben im Source (Ausgangs)- und Target (Ziel)-System durchzuführen. Dabei werden die Hierarchiestrukturen der KOS bezogen auf die Suchangaben dargelegt, um die Bedeutung und die zusammen vorkommenden Begriffe in den jeweiligen Klassen bzw. Hauptklassen zu verdeutlichen. Außerdem werden Informationen zu den jeweiligen Begriffen, wie z.B. die Registereinträge, verlinkten Mappings, Anmerkungen und synonymen Begriffe zur Suche im Zielsystem präsentiert. Durch Anklicken des *Map Buttons* wird zur gesuchten Notation der Konkordanzerstellungsprozess ausgelöst. Einen Einblick in die Konkordanzdatenbank liefert der Button *Looking up database*, woraufhin Mapping-Vorschläge im mittleren Fenster erscheinen. Die

Ergebnisse der automatischen Begriffssuche im Zielsystem und die Ergebnisse der statischen Auswertungen zusammen vorkommender Notationen in den Titeldatensätzen werden dafür ausgewertet und als Konkordanzvorschläge zur intellektuellen Überprüfung (*Mappings Candidates*) angezeigt. Das *Active Mapping*-Fenster bietet die Möglichkeit, ausgewählte Notationen bzw. Klassen zu übernehmen, zu überprüfen und lokal oder in der VZG-Konkordanzdatenbank zu speichern. Aus dem Fenster des Zielsystems können ebenfalls ermittelte Klassen zum *Active Mapping* übernommen werden.

## Literatur


- Keil, Stefan. „Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben." *Information – Wissenschaft und Praxis* 63, Nr. 1 (2012): 45–55. http://dx.doi.org/10.1515/iwp-2012-0004.

- Pfeffer, Magnus. „Automatic creation of mappings between classification systems." Vortrag auf der European Conference on Data Analysis, Luxemburg, 10. Juli 2013. Zuletzt geprüft am 05.02.2016. http://de.slideshare.net/MagnusPfeffer/pfeffer-automatic-mapping.

- Lauser, Boris, Gudrun Johannsen, Caterina Caracciolo, Johannes Keizer, Willem Robert van Hage und Philipp Mayr. „Comparing human and automatic thesaurus mapping approaches in the agricultural domain." Vortrag auf der 10th International Conference on Dublin Core and Metadata Applications, 2008. Zuletzt geprüft am 05.01.2016. http://arxiv.org/abs/0808.2246.

- Reiner, Ulrike. „Automatische DDC-Klassifizierung: Bibliografische Titeldatensätze der Deutschen Nationalbibliografie." *Dialog mit Bibliotheken* 22, Nr. 1 (2010): 23–29. Zuletzt geprüft am 05.02.2016. http://nbn-resolving.de/urn:nbn:de:101-2011012860.

- ISO. „ISO 25964–2: Information and documentation – Thesauri andinteroperability with other vocabularies. Part 2: Interoperability with other vocabularies".

- Soergel, Dagobert. „Conceptual foundations for semantic mapping and semantic search." In *Concepts in context*, herausgegeben von Felix Boteram, Winfried Gödert und Jessica Hubrich, 13–35. Würzburg: Ergon, 2011.

- Voß, Jakob.: „JSKOS data format for knowledge organization systems." Zuletzt geändert am 04.12.2015. https://gbv.github.io/jskos/.


 